

09 QoS を考慮したネットワークシステムの性能評価に関する考察

齋藤 仁[†] 今井 博英[†] 角山 正博^{††} 石井 郁夫[†] 牧野 秀夫^{†††}
[†]新潟大学大学院自然科学研究科 ^{††}新潟工科大学 ^{†††}新潟大学工学部情報工学科

1. はじめに

近年、コンピュータネットワークの普及に伴い、多くの人がリアルタイムに情報を入手することが可能になりつつある。しかし、映像や音声を含むマルチメディア情報の配送には広い伝送帯域と CPU 資源が必要となるため、品質を制御する QoS (Quality of Service) を用いたネットワークシステムが提案され、実用化されつつある。

QoS はクライアントから発生するタスクに優先度を付加し、その値に基づいてサーバが処理の順序を制御することによって実現される。しかし、負荷の変動やシステム中の故障の発生による性能低下があるほか、種々の状況の元でも常に所望の性能を維持できるシステムを実現することは容易ではない。そのため、システムを開発する際に予め性能を評価しておくことが重要になる。

本研究では QoS を用いたネットワークシステムの性能評価方法について検討する。まず、対象とするシステムの構成を示し、システムの評価のための尺度について考察する。ついで、棄却率を用いた評価を行なうためのモデル化方法を例を用いて示す。

2. 対象とするシステムの構成

本研究では、広く用いられている双方向型伝送路を持つネットワークシステムを対象とする。これを図 1 に示す。また、このシステムは QoS に基づく優先度制御がなされるクライアント/サーバシステムとして構築され、サーバは複数のプロセッサと複数の待ち行列からなっている。サーバの構成を図 2 に示す。

複数のクライアントから発生したタスクはサーバ内で優先度に基づいて処理の順序を決定された後で、プロセッサに割り当てられる。プロセッサに空きがあれば、直ちに処理が開始されるが、空きがない場合は待ち行列に入れられる。プロセッサが空き次第、優先度

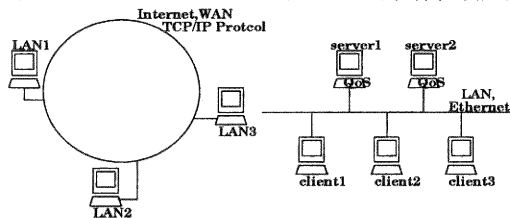


図 1: ネットワーク構成

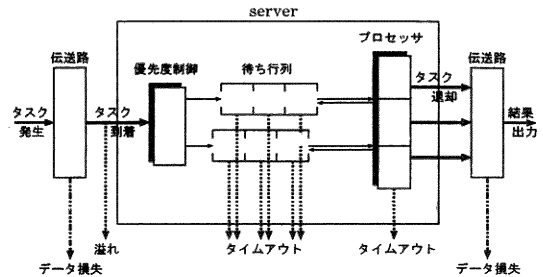


図 2: サーバの構成

の高い順に先頭のタスクから処理が開始される。また、サーバ内の待ち行列とプロセッサに割り当てられたタスクには制限時間が課せられており、制限時間を満たせないタスクはただちに捨てられる。さらに、待ち行列に空きがない時に発生したタスクは溢れとして捨てられる。伝送路内においてはコリジョン (衝突) が一定の確率で発生し、発生後は一定時間後に再送される。

3. 評価方法に関する考察

3.1 評価尺度

システムを評価する為の尺度としては、以下ののようなものが考えられる。

- 棄却率に関する尺度 … パケット、セル、タスクなどの棄却率、呼損率 (タスク要求が拒否される割合)。
- 時間に関する尺度 … 平均待ち時間、平均応答時間、収束するまでの時間、スループット。
- その他 … 配送されたマルチメディアデータに対する人間の感覚を考慮した評価など。

次に、これらの評価尺度に対応できる評価方法を検討する。

3.2 評価方法

一般化確率ペトリネット (GSPN) を用いてモデル化を行なう。この方法を用いることによって QoS を用いたシステムを比較的簡単に表現することが可能になる。さらにペトリネットをマルコフモデルに変換することによって、マルコフ連鎖法などによる解析ができる。また、GSPN はプレースとトークンの流れによって視覚的にシステムの働きを追跡することができるため、システムの振舞いを直観的に把握する事ができるという

特徴がある。本モデルを用いることによって、先に示した棄却率に関する評価、及び時間に関する評価を行なうことができる。ただし、先に示したその他の尺度を用いた評価は除外する。

4. システムのモデル化

1 サーバ2クライアントからなる QoS を用いたネットワークシステムをモデル化した例を示す. 伝送路の GSPN モデルを図 3 に, サーバの GSPN モデルを図 4 に示す.

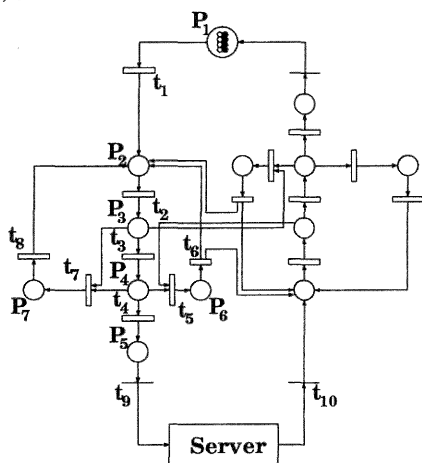


図 3: 転送路の GSPN モデル

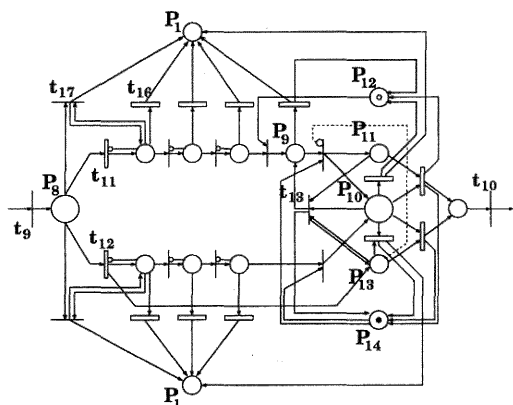


図 4: サーバの GSPN モデル

モデルを簡単にするために、システムは優先度数を2、優先度の高いタスクを●、低いタスクを○で表わす。ここで用いる伝送路では衝突を考慮する必要がある。一つのクライアントからは最初に発生したタスクの処理終了が得られない限り次のタスクは発生しないものとする、衝突は、(1) 同時に複数のクライアントがサーバへタスクを転送した場合、(2) クライアントか

らサーバヘデータの転送を開始した時、その信号が到着する前にサーバからデータの転送が開始された場合、(3) サーバからクライアントヘデータの転送を開始した時に、その信号が到着する前にクライアントから新しいタスクの転送が開始された場合、(4) サーバが同時に複数のクライアントに向けてタスクを転送しようとした場合、の四つがある。図 3 はこれらを考慮した構成になっている。

サーバには優先度の数だけの待ち行列を用意し、 P_8 において優先度に基づいた待ち行列に振り分けられる。 P_{10} はプロセッサにタスクが割り当てられているかどうかを示し、低優先度のタスク○を処理中に高優先度のタスク●が発生すると、 t_{13} が発火し、処理中の低優先度タスク○は待ち行列の先頭を示す P_9 に退避させられる。代わりに高優先度のタスク●の処理が開始される。待ち行列やプロセッサに課せられた制限時間を過ぎると、 t_{16} などのトランジションが発火し、タイムアウト処理が行なわれる。さらに待ち行列に空きがない場合には t_{17} が発火し、溢れ処理がなされる。図4はこれらの処理を考慮したサーバの構成になっている。

5. まとめ

QoSを考慮したネットワークシステムの性能評価を行なうためのシステムのモデル化について考察を行なった。ここでは、性能評価の尺度として棄却率に関する尺度、及び時間に関する尺度を用いたGSPNによるモデル化方法を示した。GSPNを用いてモデル化することにより、複雑なQoSネットワークシステムを比較的簡単にモデル化することができる。今後の予定としては、以下のような事があげられる。

- 今回モデル化したシステムの棄却率と時間に関する尺度に基づく解析方法の検討
- カラーペトリネットの解析

参考文献

- [1] Oliver C.Ibe, Hoon Choi, Kishor S.Trivedi : "Performance Evaluation of Client-Server Systems", IEEE Transactions on Parallel and Distributed Systems, **4**, 11, pp1217-1229, November 1993.
- [2] Jogesh K.Muppala, Steven P.Woolet, Kishor S.Trivedi : "Real-Time-Systems Performance in the Presence of Failures", IEEE Computer, **24**, 5, pp37-47, May 1991.